Ramesh C. Jain Anil K. Jain
Editors

Analysis and Interpretation of Range Images

With 161 Illustrations

Springer-Verlag
New York Berlin Heidelberg
London Paris Tokyo Hong Kong

Ramesh C. Jain
Electrical Engineering and
Computer Science Department
University of Michigan
Ann Arbor, MI 48109
USA

Anil K. Jain
Department of Computer Science
Michigan State University
East Lansing, MI 48824
USA

Series Editor

Ramesh C. Jain
Electrical Engineering and
Computer Science Department
University of Michigan
Ann Arbor, MI 48109
USA

Printed on acid-free paper.

Camera-ready copy provided by the authors using LaTeX.
Printed and bound by Edwards Brothers, Inc., Ann Arbor, Michigan.
Printed in the United States of America.

9 8 7 6 5 4 3 2 1

ISBN 0-387-97200-5 Springer-Verlag New York Berlin Heidelberg
ISBN 3-540-97200-5 Springer-Verlag Berlin Heidelberg New York

4

Segmentation versus object representation – are they separable?

Ruzena Bajcsy
Franc Solina
Alok Gupta[1]

4.1 Introduction

When vision is used for moving through the environment, for manipulating or for recognizing objects, it has to simplify the visual input to the level that is required for the specific task. To simplify means to partition images into entities that correspond to individual regions, objects and parts in the real world and to describe those entities only in detail sufficient for performing a required task. For visual discrimination, shape is probably the most important property. After all, line drawings of scenes and objects are usually sufficient for description and subsequent recognition. In computer vision literature this partitioning of images and description of individual parts is called segmentation and shape representation. Segmentation and shape representation appear to be distinct problems and are treated as such in most computer vision systems. In this paper we try to disperse this notion and show that there is no clear division between segmentation and shape representation. Solving any one of those two problems separately is very difficult. On the other hand, if any one of the two problems is solved first, the other one becomes much easier. For example, if the image is correctly divided into parts, the subsequent shape description of those parts gets easier. The opposite is also true when the shapes of parts are known, the

[1]GRASP Laboratory,Computer and Information Science Department, University of Pennsylvania, Philadelphia, PA 19104, USA. This work was supported in part by the following contracts and grants: NSF DCR-84-10771, NSF ECS-84-11879 and DMC-85-12838, US Postal Service contract 104230-87-H-0001/M-0195, Air Force F49620-85-K-0018, F33615-83-C-3000, F33615-86-C-3610, DARPA/ONR grants NOO14-85-K-0807, NOOO14-85-K-0018, ARMY DAAG29-84-K-0061 and DAAG29-84-9-0027, NSF-CER MCS-82-19196, DCR-82-19196 A02, NIH NS-10939-11 as part of the Cerebrovascular Research Center, by DEC Corp., and the LORD Corp.

partitioning of the image gets simpler. Since neither of them can be easily solved in isolation, at least not on the first try, we argue that they should interact to guide and correct each other. Hence, segmentation and shape recovery should not be studied separately. The complete visual interpretation problem is even more complex because the initial data acquisition process should not be separated from the later segmentation and shape representation. How data acquisition can interact with the interpretation stage is investigated in computer vision under the heading of active vision [Baj85]. In this paper we concentrate only on the interaction between segmentation and shape representation, assuming an image taken from a particular viewpoint is given.

A more formal problem definition of the topic of this paper is the following. Given an arbitrary spatial arrangement of static, three dimensional solids, imaged by a noncontact sensor, answer the following three questions:

1. What are the geometric primitives that (possibly uniquely) describe the data?

2. What are the processes that carry out this decomposition?

3. What is the overall control strategy to explain the measured data?

While the first two questions represent the analysis aspect of the problem, the last one can be explained as the synthesis or integration of the whole system.

In the rest of the paper we assume that a complete depth map of a scene is given. Obtaining a depth map is one of the stated goals of low level vision modules, such as stereo and shape from shading. The computation of the depth map or 2-1/2 D sketch was once considered to be the harder part and that image interpretation from there on would be easy. Although dense and accurate depth maps are now available from laser range scanners, the interpretation of those images is still difficult. A depth map as the starting point, obtained either with a laser scanner or from low level image techniques on gray level images, does not simplify neither segmentation nor shape recovery in any large extent. For the examples in this paper we use range images taken from a single viewpoint [Tsi87]. Due to self occlusion, not all points on the surface of an object are given. Since the supporting surface is fixed, range points from the support can be easily removed at the start of scene interpretation.

When the necessity for interaction between segmentation and shape representation is acknowledged, control strategies that implement this strategy in a vision system become important. The influencing factors on the design of the control strategy are the goal of the vision system, the scene complexity and the dimensionality of the objects in the scene. Typical goals of a vision system are locating obstacles in a scene for mobile robot navigation, enabling manipulation with robot hands or identifying objects by

matching recovered shape descriptions to a given data base. The complexity problem is to find out whether the scene contains a single convex object, a non-convex object consisting of parts, or more than one object. Scene classification according to its complexity can greatly simplify the control structure for interpretation. Establishing dimensionality is to find out if a scene can be interpreted only in terms of volumetric models, flat-like models or rod-like models. Global measures such as center of gravity and moments of inertia give such estimates. The importance of dimensionality parameters is that, depending on the dimensionality, different geometric primitives come into play. For example, in the case of a scene with flat-like objects only, surface primitives should be sufficient and no volumetric primitives would be required. A segmentation system for intensity images that uses such adaptable parameters, provided by the user and computed from the image data, is described in Anderson et al [ABM88].

Depending on all those influencing factors, different geometric parameters can be used for shape discrimination to recover volumetric, surface or boundary properties. One of the hardest problem that the computer vision community has tried to solve during the last 20 years is the extraction of geometric shape properties. The rest of the paper is organized as follows: problems and issues in selecting the type of shape primitives are in section 4.2, section 4.3 is on segmentation, and section 4.4 on the overall control structure. In section 4.5 we compare the actual occluding boundaries of objects in range images to the boundaries of volumetric models fitted to the data to point out the different scope of those models. Section 4.6 is a summary.

4.2 The Role of Shape Primitives

Decomposition into parts, units or primitives is the basis of scientific methodology. Because of the limits on how much information we can process at a time, we have to simplify and view the world at various levels of abstraction. In shape decomposition, one tries to follow the principle of orderliness, which means partitioning things in the simplest possible way. Such partitioning normally reflects the structure of the physical world quite well due to the principle of parsimony [Arn74]. The choice of primitives can be guided by some general requirements such as a unique decomposition into primitives, that the primitives cannot be further decomposed or that the set of primitives is complete. Some of the shape representation criteria are designed primarily to facilitate object recognition when models recovered from images are matched to a model data base. For a discussion of different criteria for shape representation we refer the reader to [Bra83]. Unfortunately, all those principles have not been applied to any general shape representation scheme for 3-D objects. A review of computer vision literature which reveals the large variety of geometrical primitives

that were investigated for their applicability to shape representation is a testimony to the difficulty of shape description [BJ85b]. Another discipline involved in representing shape is computer graphics, but from a synthesis (generating) point of view. Some commonly used 3-D representations in graphics are wire-frame representation, constructive solid geometry representation, spatial-occupancy representation, voxel representation, octree representation, and different surface patch representations. Splines are used for surface boundary representation.

In early days of computer vision, most shape primitives were borrowed from computer graphics. But requirements for shape primitives in computer vision are different from the ones for computer graphics. Foremost, shape primitives for computer vision must enable the analysis (decomposition) of shape. Common shape primitives for volume representation are polyhedra, spheres, generalized cylinders, and parametric representations such as superquadrics. Different orders of surface patches (planar, quadratic, cubic) are used for surface representation. For boundary description one can use linear, circular or other second order models for piecewise approximation, and higher order spline descriptions. In the rest of this section we will discuss what influences the selection of shape primitives in computer vision.

If only one shape primitive is chosen, the segmentation process is relatively simple. But the resulting segmentation may not be natural! The data can be artificially chopped into pieces to match the primitives. An example of such unnatural decomposition is when a circle is represented piecewise with straight lines or when a straight line is represented with circular segments. If the scene consists of both straight lines and circles, then neither straight lines nor circles alone would enable a natural segmentation. A natural segmentation, on the other hand, would partition an image into entities that correspond to physically distinct parts in the real world. A solution to such problems is to use more primitives. How many primitives are required for segmentation of more complicated, natural scenes is then the crucial question. The larger the number of primitives, the more natural and accurate shape description and segmentation is possible. But the larger the number of primitives, the more complicated gets the segmentation process. Finding the right primitive to match to the right part of the scene leads potentially to a combinatorial explosion. This argues for limiting the number of different shape models.

Another influencing factor on the number of different models is the level or granularity of models. A large number of low level models is required for scene description because of their small size or granularity. Low level models can fit to a large variety of data sets but bring little prior information to the problem. Hager [Hag88] calls low level models descriptive as opposed to prescriptive and are as such used mostly in data-driven vision systems. Substantial manipulation is required to obtain further interpretation of the data by aggregating low level models into models of larger granularity

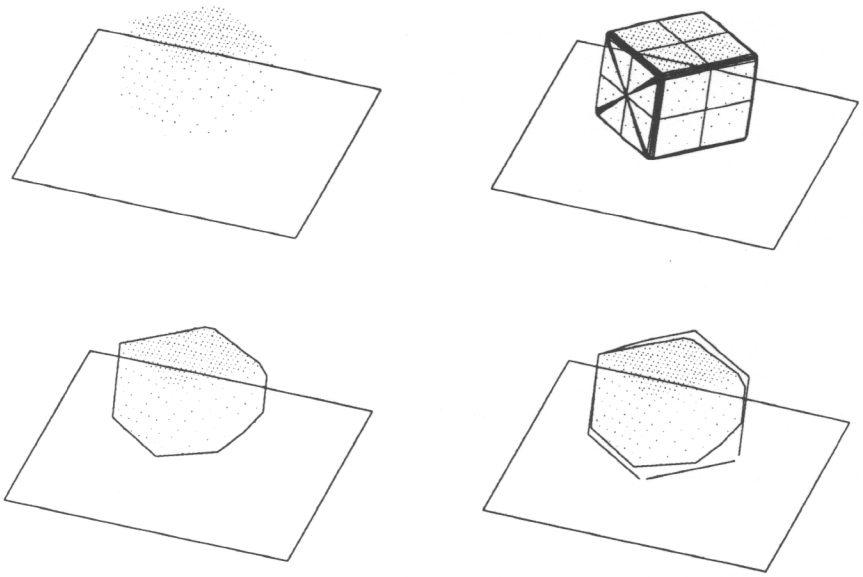

FIGURE 4.1. Top left is a laser range image of a cube. Top right is a superquadric volumetric model recovered by least-squares fitting of a parametric function to the 3-D points. This model gives probably the best overall explanation of the imaged object, an explanation, that most human observers would likely agree with. A closer examination by following the occluding contour reveals, however, that some points are missing from the lower right corner of the cube. Bottom left is a line approximation that closely follows the local shape of the occluding boundary. From this edge model alone is quite difficult to conclude that the object is roughly shaped as a cube. This difficulty is due to the fairly low level of line models which bring very little prior information to the interpretation problem. Volumetric models, on the other hand, are of larger granularity. By bringing additional information in form of internal parametrization such volumetric models constrain the recovery problem and find a plausible solution.

which correspond to real world entities. Such aggregation techniques often fail because it is not possible to distinguish data from noise or account for missing data only on the basis of local information. Higher level models, on the other hand, are prescriptive in the sense that they bring in more constraints and provide more data compression. But higher level models might miss some important features because they cannot encompass those data variations within their parametrization. A concise model which adequately describes the data will enable partitioning or segmentation of images into right parts and ignore noise and details beyond the scope of the task.

In everyday life, people use most of the time a default level of representation, called basic categories [Ros78]. Basic categories seem to follow natural breaks in the structure of the world which is determined by part configu-

ration [TH84]. Shape representation on the part level is then very suitable for reasoning about the objects and their relations in a scene. For part level description in vision, a vocabulary of a limited number of qualitatively different shape primitives [Bie85] and different parametric shape models have been proposed. Parametric models describe the differences between parts by changing the internal model parameters. In computer vision, the most well known parametric models suitable for representing parts are generalized cylinders but superquadrics with global deformations seem to have some important advantages when it comes to model recovery [BS87,Pen86]. This is discussed later in this section.

It is sometimes possible to know a priori that a certain class of geometric models is sufficient to describe observed data. Another possibility is to somehow evaluate the complexity of the scene and the dimensionality of the objects in the scene. Knowing the complexity of the scene can greatly simplify the control structure for segmentation and shape recovery while knowing the dimensionality of objects simplifies the selection of shape models.

The objective of a vision system, whether the goal is to avoid obstacles during navigation, to manipulate objects with robotic grippers and hands or to identify objects by matching them to a data base, is another constraint during shape model selection. For object avoidance, only representation of occupied space is necessary, often allowing to largely overestimate the size of obstacles. In addition to location and orientation, grasp planning for robotic hands requires knowing more precisely the size and overall global shape of the object. For object recognition, more specific, identifying features are needed. Different shape primitives are better at representing different aspects of shape and at different scales. Volumetric representation provides information on integral properties, such as overall shape, enabling classification into elongated, flat, round, tapered, bent, and twisted primitives. They can best capture the overall size and volume since they must make an implicit assumption about the shape of the object hidden by self occlusion. Surface representation is better at describing details that pertain to individual surfaces which can be part of larger volumetric primitives. Surface primitives can differentiate planar surfaces versus curved surfaces, concave versus convex, and smooth versus undulated surfaces. Boundary representation lies, in scope, somewhere between volumetric and surface representation. On one hand, it is a local representation of curvature and surface near the boundaries, on the other hand, by delineating the boundaries of an object from the background, it defines the whole object.

Coming back to the problem of segmentation, which is to match the right kind of shape model with the right parts of data in an image, brings up the question of facilitating this matching process. Instead of a combinatorial search, one should find a way of determining from the data, which models to use where. A possible way to cut the search is by using a coarse to fine strategy. To find such a shortcut leads back to the question in the title of

this paper is segmentation separable from shape representation? By now it should be clear that the segmentation process and its results depend on the selected shape primitives. To facilitate segmentation we believe that for a general purpose vision system one needs volumetric, surface and boundary shape primitives.

We now outline the criteria for selection of shape primitives, in particular the superquadrics family of shapes as volumetric primitive. As mentioned before, complexity and dimensionality of the scene determine the shape primitives that will adequately describe the scene. The primitives should be able to define local as well as global characteristics of the parts of objects in the scene. In a general scene objects can have parts and parts themselves consist of parts. In this paper we are addressing the issue of segmenting parts of an object in the scene. The problem of segmenting individual objects in a complex scene can be handled as an extension of current approach. The object shape can be described at three levels of complexity, each contributing to the overall shape :

1. Description in three dimensions using superquadrics shape primitive.

2. Description in two dimensions using surface primitives.

3. Description in one dimension using contour primitive.

Low level models like contours and edges have low granularity (see figure 4.1) and are too local to capture or make use of the gross structure of the world. They are sensitive to local changes and difficult to put together in a global context. However, this characteristic allows them to capture local details of shape that would be missed or smoothed out by more global primitives. When analyzed as a whole, contour primitives have the remarkable capability of describing global shape and segmenting an object into parts, as demonstrated in [HR85].

The next level of shape description is achieved by describing local and overall surface characteristics. Binford [Bin82] has argued that parts should be defined by continuity. Human perception defines objects as collection of surfaces and does not necessarily segment objects at surface discontinuities. Nonetheless, surfaces play important role in human perception of shape. A lot of effort in computer vision has been spent on describing complex surfaces as piecewise continuous patches. In order to arrive at a global interpretation, a surface representation scheme that combines relevant surface characteristics (e.g. lines of curvature) with the surface patches is needed. This would enable us to describe object shape at a higher level (e.g, The surface has 3 pleats) than that described by individual patches.

Parametric models like generalized cylinders and their derivatives have been used as volumetric primitives by vision researchers because they give compact overconstrained estimate of overall shape. This overconstraint comes from using models defined by a few parameters to describe a large

set of 3-D points. Researchers have developed rule-based systems to recover generalized cylinders from image data. In such systems monitoring of progress is difficult and a direct evaluation criteria of results is not available. Also, they can recover only a restricted subset of generalized cylinders, such as linear straight homogeneous generalized cylinders [NB77]. Superquadric models use least squares minimization for recovery of their parameters. An important advantage for ease of model recovery is that the superquadric surface is defined by an analytic function, differentiable everywhere. Superquadric shapes form a subclass of shapes describable by generalized cylinders. Shape deformations like bending and tapering can be defined with global parametric deformations. Superquadrics with parametric deformations encompass a large variety of natural shapes yet are simple enough to be solved for their parameters. Due to their built-in symmetry, superquadric models predict the shape of occluded parts conforming with the principle of parsimony - among several hypotheses select the simplest [Gom62].

If one accepts the need for multiple representations, one has to have a control strategy to bring all of them together. But first, one has to decide what is needed for segmentation.

4.3 Segmentation Process

There are two basic strategies for segmentation:

1. Proceed from coarse to fine discrimination by partitioning larger entities into smaller.

2. Start with local models and aggregate them into larger ones.

Both of these strategies have been used in the past [DH73,Pav77]. The advantage of the coarse to fine strategy is that one gets first a quick estimate about the volume/boundary/surface of the object which can be further refined under control of some higher level process which determines how much details on wishes to know. The disadvantage of this approach is that the amount of detectable detail is not always sufficient without switching to a different kind of representation. For example, to describe smaller shape details one might have to go from volumetric to surface representation. This progression of looking at data at different scales is more formalized in scale-space [Wit84] and in different multiresolution signal decomposition techniques [Mal88]. The important idea that these methods convey is that progressive blurring of images clarifies their deep structure [KvD79]. Large scale structure constrains the structure at finer levels so that adding details only entails adding information and does not require changing the larger structure. Although these multiresolution techniques do not correspond to structural decomposition of images into parts, one assumes that the

same principle applies there, also. When a part model must be subdivided into smaller parts to gain finer resolution it should not affect the original partitioning. In that sense, backtracking to change prior decisions would not be necessary.

The second strategy, which goes from local to global, starts with local features and incrementally builds larger representations. This can be an advantage or disadvantage at the same time. Some details could help the classification process early on by excluding any hypothesis that clearly does not include such particular details. On the other hand, keeping track of too many details at once can lead to a combinatorial explosion. As already mentioned, aggregation of low level models into models of larger granularity is difficult in presence of noise or when data is missing. It is also necessary to ignore details that cannot be represented in the next higher level of representation. Recovering from mistakes or erroneous aggregations by rearranging the low level models in new ways should be possible.

Both methods of segmentation, top-down and bottom-up have their benefits and problems. Both methods should be used in a general vision system and the question is how to combine them in a fruitful way. Another possible way of dividing segmentation methods is by the type of shape primitives they use. The following three subsections are on segmentation using volumetric, boundary, and surface representation.

4.3.1 Segmentation using volumetric representation

Although many different methods for partitioning into volumetric primitives exist, we shall focus only on two examples that typify such use of volumetric primitives. The first one is the work by Binford and Nevatia [NB77] who used generalized cylinders for describing parts of objects. They start from local edge models, cross sections and aggregate them into parts, each of them represented with a generalized cylinder. Many improvements of this basic method exist, both by the original authors and others ACRONYM being probably the most well known system [Bro83]. This is an example of a strategy going from local to global aggregates.

An example of the global to local method of segmentation is the superquadric fitting method by Solina [Sol87]. Here the goal is to decompose objects or scenes into parts which can be represented with a single superquadric model enhanced with global deformations such as tapering and bending. Since a superquadric surface can be described with an analytic function, an iterative least-squares minimization of a fitting function can be used for shape recovery. Consider a depth map of an arbitrary scene. The initial model is an ellipsoid in the right position, orientation and of the right size to cover all of the 3-D points. During the least-squares minimization, the shape of the initial model starts to change so that the given

range points would lie on or close to the surface of the model. If the model can reject and accept 3-D points, the model can actively search for a better fit, resulting in a recursive subdivision of the scene into parts. The simplest case is when only a single part is present in the scene. Then the model must incorporate all of the points. When several parts or objects made up of multiple parts are present, a suitable distance measure must be used to decide which 3-D points should be included in a particular volumetric model and which points should be excluded. This question has not yet been successfully solved. The same problem of sensitivity and robustness, however, is present in the aggregation method where the setting of the similarity parameters for joining features into larger entities must be robust enough to bridge small gaps in measurements due to noise and imperfect fit, yet sensitive enough to distinguish between different parts.

Given some complexity measures for the scene, the segmentation process can be changed accordingly. In the one-object scenario one can first fit a volumetric model and then analyze how well the model fits the data and adjust the shape and deformation parameters for a better fit. If several objects are present, one should apply segmentation to each cluster individually. In the difficult case, when a heap of objects is given with multiple occlusions, one might concentrate only on the top most object and treat it in the same way as in the one object scenario.

4.3.2 Segmentation using boundary information

The segmentation process using boundary information is based on the detection of discontinuities both in depth values as well as in orientation. Given discontinuities in depth and orientation, similar adjoining segments can be merged and curve fitting, using splines or some other piecewise model can be performed. Partitioning that corresponds to the human notion of parts can be achieved using changes in curvature of the occluding boundary to detect concavities which indicate part boundaries [HR85]. Occluding contours play a large role in human perception. Strong spatial impressions arise from seeing only silhouettes of objects in a general orientation. Koenderink relates this to the capability of inferring from occluding boundaries the shape of the near lying surface [Koe84]. Ramachandran [Ram88] shows how boundaries influence also the interpretation of shaded surfaces. When information from shading underdetermines the interpretation, information from borders helps to resolve ambiguity throughout the image.

4.3.3 Segmentation using surface primitives

A large portion of computer vision literature is on different methods for surface reconstruction. A recent overview of different surface reconstruction approaches can be found in [BZ87a]. The reason for the widespread

interest in surface reconstruction is that this fits well into the prevalent bottom-up approach in vision and that surface is a much more tangible property than volume. Surface segmentation can be based either on merging similar local surface models, or by defining region boundaries in terms of differential geometry [BJ86]. The aggregation process begins with small local neighborhoods which are then combined if they are similar in depth values, surface normal values or some curvature measurements. The result is a scene segmented into surface regions with similar surface characteristics. The difficulty with both surface segmentation approaches is that it is sensitive to local variations which are not important but are difficult to eliminate unless the larger context is taken into account. Since this larger context can be much easier accounted for by volumetric models, it should be here where the surface, volume and boundary segmentation could cooperate. We have implemented such segmentation process in Gupta [Gup88].

4.4 Control Structure

The problem that we wish to address in this section can be stated in the following way. Given that we have all three different modules for extracting volume, surface and boundary properties, how should they be invoked, evaluated and integrated? There are two extreme possibilities. The first one is to apply all three modules simultaneously. The second is to apply them strictly in a predetermined sequence. In the parallel approach conflicting hypothesis can arise that would have to be resolved. The sequential method may lead the segmentation process in a wrong direction so that backtracking would sometimes be necessary. A combined approach where all three methods could interact would not be so vulnerable. This opens up the problem of evaluating and comparing information embedded in models built by different aggregation methods. What do you do if different types of models do not mutually reinforce each other? In such cases, one would normally prefer models of smaller granularity that are less prescriptive models that closely follow the data in the image . But this has to be distinguished from the case when the information that could give rise to low level models is not present. A good example are the well known phenomena of illusory contours in human perception. We can perceive solid shapes although a large part of boundary lines physically do not exist. In conflicting situations information has to be reorganized and the control system adapted. Anderson et al [ABM88] designed an adaptive system for 2-D segmentation of intensity images based on on the general assumption that the gradient value at region borders exceeds the gradient within regions. An adaptive control system that has to reconcile conflicting shape models might use also some result from the recent study of active reduction of uncertainty in multi-sensory systems [Hag88].

To incorporate the best of the coarse to fine and fine to coarse segmen-

tation strategy we propose to perform volume and boundary fitting in parallel, followed by surface description. The volumetric shape recovery that we have in mind is a global, holistic method going from very coarse to fine fitting on the part level while boundary detection and description which is local by the nature of the data can guide segmentation. These two processes are complementary in the approach of explaining the data, accounting for global position, orientation, size and shape such that the local boundary confirms with the boundary obtained from the volumetric fitting. Surface modeling is necessary for representing details that cannot be encompassed by part-level volumetric models. Surface fitting can be used also to reaffirm segmentation into parts by testing the surface continuity or discontinuity between parts.

The control structure has to determine the reliability of information obtained from each primitive. Superquadrics being part-models, need to be compared with the bounding contour and available surface points to evaluate suitability of the recovered model. Surfaces, for most part, complement the information provided by bounding contours. Bounding contours are viewpoint dependent and may not account for all relevant contours needed for complete segmentation or description. This is obviously the case when viewpoint is not general. Thus, in some cases, surface information along with bounding contour can determine if the object is in most general position or not and ask for information from different viewpoint (or rotate the object). For some objects, it may not possible to obtain data from a viewpoint such that the object can be segmented by analyzing only the contour. In such a case, if surface information strongly suggests segmentation along a surface discontinuity, bounding contour should not lower our confidence in surface information. On the other hand, if contour suggests a possible segmentation and there is no support from surfaces, a decision will have to be made about the possibility of segmentation assuming a smooth join between part and object body.

Superquadrics essentially provide global description of individual parts and give the feedback as to the possibility of further segmentation of that part. They lack the local information needed to suggest possible segmentation sites. Contour and Surfaces, on the other hand, actively hypothesize and carry out segmentation. The process continues until a satisfactory description of parts is achieved.

During the segmentation process the control module has also to decide on part/whole (or part/detail) relationships. This requires determining the scale of a potential part given the overall size of the object and deciding to consider it a part or just a detail of the object that can be ignored (implying that current description is adequate).

The global control program must have many parameters and thresholds that would have to be predetermined or, if possible, adjusted during the process. Some of those parameters are the following:

- the size (or range of sizes) of the local neighborhood for local processing,
- the size (or range of sizes) of volumetric models,
- the number (or range) of expected segmented units,
- all the thresholds (for partitioning and aggregation),
- the level of details that we wish to explain.

4.5 Results

We applied the volumetric shape recovery procedure [Sol87] to a set of range images of single objects (Figures 4.2, 4.3 and 4.4). The contour obtained by tracking the occluding boundary and the contour of the recovered volumetric model are compared in all cases. While the volumetric model gives a holistic explanation of the whole object it can miss details that are beyond the scope of the model. An overall measure of goodness of fit, like the residual from least-squares fit [Sol87], does not always give an accurate evaluation of the appropriateness of the volumetric model. Although models can have about the same overall goodness of fit, like the volumetric models in Figures 4.2 and 4.3, they can be more or less acceptable representations of the actual object. Comparing the local boundary of range points with the boundary of the recovered volumetric model can point out the aberrations of the volumetric model and suggest improvements in segmentation or refinement in shape representation. When boundaries do not coincide, preference should be given to actual boundary in the range image, but the possibility of missing data (i.e. occlusion) must be considered also. For example, the actual occluding boundary in Figure 4.2 is without doubt a better representation of the object while the actual boundary shown in Figure 4.1 probably differs from the boundary of the volumetric model because of missing range data.

4.6 Summary

In this paper we discuss some general issues concerning shape representation and segmentation in computer vision. The selection of shape models should be guided by the task of the vision system, the complexity of the scene and the dimensionality of objects in the scene. We argue that shape representation and segmentation should not be approached separately. By picking a particular shape model we restrict the possible ways of partitioning or segmenting an image. Volumetric, boundary and surface models

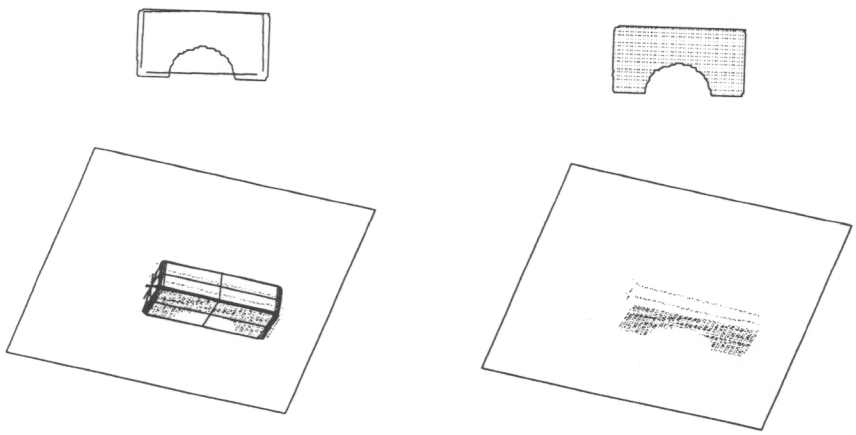

FIGURE 4.2. Range image of a block with a circular cutout. Top left is the original range image. Top right is the best fitting volumetric model. Bottom left is a the line-approximation of the occluding contour as seen from top. Bottom right is the comparison of the occluding boundary with the boundaries of the volumetric model from above. The circular cutout was not accounted for by the volumetric model.

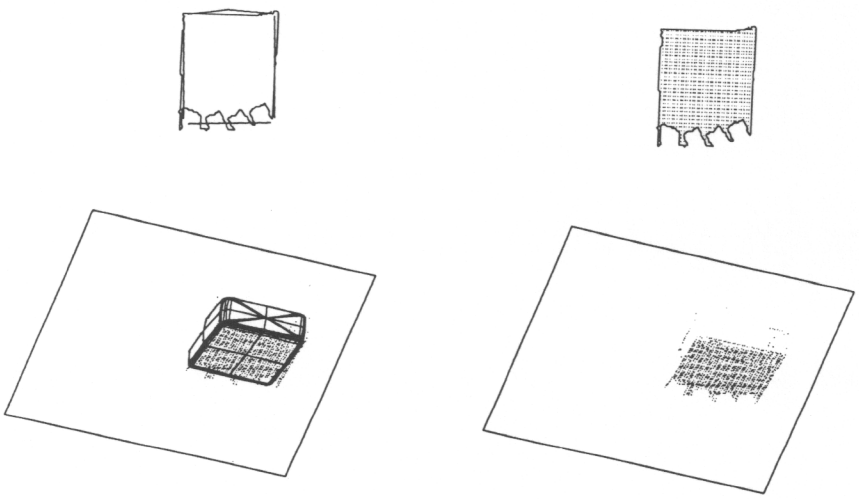

FIGURE 4.3. Range image of a block with a jagged edge. Top left is the original range image. Top right is the best fitting volumetric model. Bottom left is the line-approximation of the occluding contour as seen from top. Bottom right is the comparison of the occluding boundary with the boundaries of the top volumetric model. Since the differences between the two outlines are small in comparison with the overall size of the object the jagged edge could be brushed away as a detail.

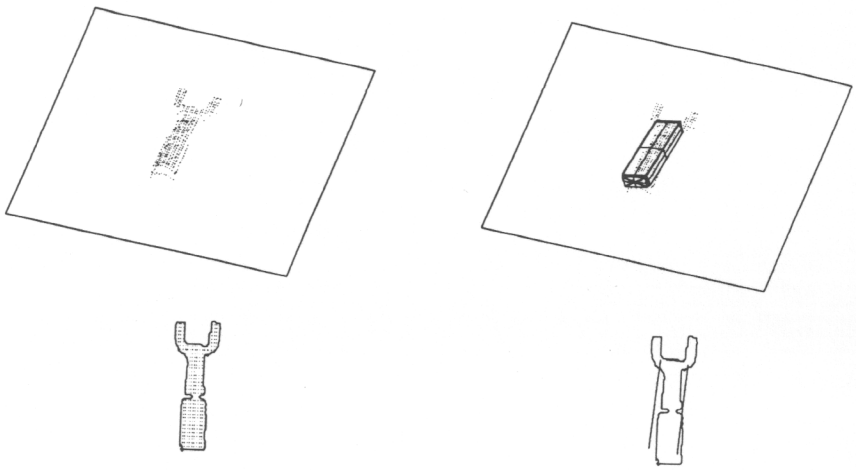

FIGURE 4.4. Top left is the original range image of a wrench. Top right is the best fitting volumetric model. Bottom left is the line-approximation of the occluding contour as seen from top. Bottom right is the comparison of the occluding boundary with the boundaries of the top volumetric model. The two boundaries coincide only in part of the image alerting to the fact that the object consists of more than one part.

represent different types of features and at a different scale. In a general vision system, all three types of shape models should be used. We propose a control structure for such a general system which follows a coarse to fine strategy. It starts with recovery of volumetric models, constrained by occluding contours for segmentation. In order to describe finer details that cannot be encompassed with volumetric models, one has to switch to surface representation. We show some examples of how comparing the occluding boundaries can guide or correct the recovery of volumetric models. In the discussion of control structure, we stress the importance of checking not only the global goodness of fit of the applied shape models but also the local alignment in order to correct or refine the representation. The control system should also adapt to different task requirements and complexities of the scene.